

00862.022537



#2  
rec  
6-26-02

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: )  
Tetsuo KOSAKA, ET AL. ) Examiner: Unassigned  
Application No.: 10/086,740 ) Group Art Unit: 2641  
Filed: March 4, 2002 )  
For: SPEECH RECOGNITION SYSTEM ) June 21, 2002  
AND METHOD, AND )  
INFORMATION PROCESSING )  
APPARATUS AND METHOD USED )  
IN THAT SYSTEM )

RECEIVED

JUN 24 2002

Technology Center 2600

Commissioner for Patents  
Washington, D.C. 20231

SUBMISSION OF PRIORITY DOCUMENT

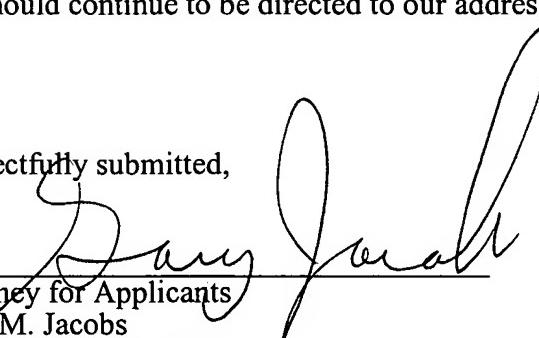
Sir:

In support of Applicants claim for priority under 35 U.S.C. § 119, enclosed is a certified copy of the following foreign application:

2001-065383, filed March 8, 2001.

Applicants' undersigned attorney may be reached in our Washington, D.C. office by telephone at (202) 530-1010. All correspondence should continue to be directed to our address given below.

Respectfully submitted,

  
Attorney for Applicants  
Gary M. Jacobs  
Registration No. 28,861

FITZPATRICK, CELLA, HARPER & SCINTO  
30 Rockefeller Plaza  
New York, New York 10112-3801  
Facsimile: (212) 218-2200  
GMJ/lmj

10/086

TETSUO KOSAKA, ET AL.

"SPEECH RECOGNITION SYSTEM AND METHOD, AND  
INFORMATION PROCESSING APPARATUS AND METHOD  
USED IN THAT SYSTEM"

G.H.O. 2641 (translation of the front page of the priority document of  
Japanese Patent Application No. 2001-065383)



RECEIVED

JUN 24 2002

Technology Center 2600

This is to certify that the annexed is a true copy of the  
following application as filed with this Office.

Date of Application: March 8, 2001

Application Number : Patent Application 2001-065383

[ST.10/C] : [JP 2001-065383]

Applicant(s) : Canon Kabushiki Kaisha

March 29, 2002

Commissioner,

Japan Patent Office

Kouzo OIKAWA

Certification Number 2002-3022047

10/086,740

TETSUO KOSAKA, ET AL

"SPEECH RECOGNITION

INFORMATION PROCESSING

APPLICATIONS AND METHODS

USING A TNAF SYSTEM"

Q.A.U. 2641

JUN 21 2002

SYSTEM AND METHOD, AND

本 国 特 許 庁

JAPAN PATENT OFFICE

CFM 2537 US

INFORMATION PROCESSING

APPLICATIONS AND METHODS

USING A TNAF SYSTEM"

Q.A.U. 2641

別紙添付の書類に記載されている事項は下記の出願書類に記載されて  
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed  
with this Office

出願年月日

Date of Application:

2001年 3月 8日

RECEIVED

JUN 24 2002

出願番号

Application Number:

特願 2001-065383

[ST.10/C]:

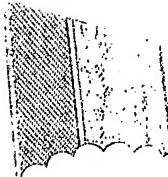
[JP 2001-065383]

Technology Center 2600

出願人

Applicant(s):

キヤノン株式会社



CERTIFIED COPY OF  
PRIORITY DOCUMENT

2002年 3月 29日

特許庁長官  
Commissioner,  
Japan Patent Office

及川耕三



出証番号 出証特 2002-3022047

【書類名】 特許願

【整理番号】 4413114

【提出日】 平成13年 3月 8日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 3/00

【発明の名称】 音声認識システム及び方法及び該システムに用いる情報  
処理装置とその方法

【請求項の数】 27

【発明者】

【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会  
社内

【氏名】 小坂 哲夫

【発明者】

【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会  
社内

【氏名】 山本 寛樹

【特許出願人】

【識別番号】 000001007

【氏名又は名称】 キヤノン株式会社

【代理人】

【識別番号】 100076428

【弁理士】

【氏名又は名称】 大塚 康徳

【電話番号】 03-5276-3241

【選任した代理人】

【識別番号】 100112508

【弁理士】

【氏名又は名称】 高柳 司郎

【電話番号】 03-5276-3241

【選任した代理人】

【識別番号】 100115071

【弁理士】

【氏名又は名称】 大塚 康弘

【電話番号】 03-5276-3241

【選任した代理人】

【識別番号】 100116894

【弁理士】

【氏名又は名称】 木村 秀二

【電話番号】 03-5276-3241

【手数料の表示】

【予納台帳番号】 003458

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0102485

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声認識システム及び方法及び該システムに用いる情報処理装置とその方法

【特許請求の範囲】

【請求項 1】 音響情報を入力する入力手段と、

前記入力手段で入力した音響情報を分析して特徴量パラメータを取得する分析手段と、

前記分析手段で得られた特徴量パラメータに基づいて圧縮符号化のための処理条件を設定する設定手段と、

前記分析手段で得られた特徴量パラメータを前記処理条件に従って圧縮符号化する変換手段と、

前記設定手段で設定された処理条件と、前記変換手段で圧縮符号化された特徴量パラメータに基づいて音声認識をする認識手段と

を備えることを特徴とする音声認識システム。

【請求項 2】 前記分析手段と前記設定手段と前記変換手段とを有する第1の装置と、前記認識手段を有する第2の装置とで構成され、

前記設定手段で設定された設定条件と前記変換手段で取得されたデータを前記第1の装置から前記第2の装置へ通知する通知手段を更に備えることを特徴とする請求項1に記載の音声認識システム。

【請求項 3】 前記認識手段は、

前記処理条件を参照して前記圧縮符号化された特徴量パラメータを復号する復号手段を備え、

前記復号手段で復号された特徴量パラメータに基づいて音声認識処理を実行することを特徴とする請求項1または2に記載の音声認識システム。

【請求項 4】 前記第2の装置が、前記通知手段で通知された前記処理条件を保持する保持手段を更に備えることを特徴とする請求項2に記載の音声認識システム。

【請求項 5】 前記認識手段は、

前記処理条件と音響モデルに基づいて音声認識に関わる尤度計算の一部を行な

う計算手段と、

前記計算手段による計算結果を用いて、前記変換手段で取得されたデータに対する尤度計算を行ない音声認識結果を得る手段とを備えることを特徴とする請求項1または2に記載の音声認識システム。

【請求項6】 前記計算手段で算出された計算結果を保持する保持手段を更に備えることを特徴とする請求項5に記載の音声認識システム。

【請求項7】 前記変換手段は、前記分析手段により得られた多次元音声パラメータを次元毎にスカラ量子化することを特徴とする請求項1乃至6のいずれかに記載の音声認識システム。

【請求項8】 前記スカラ量子化において、LBGアルゴリズムを用いることを特徴とする請求項7に記載の音声認識システム。

【請求項9】 前記スカラ量子化において、量子化対象データがガウス分布することを仮定し、量子化ステップが該分布で等確率となるよう量子することを特徴とする請求項7に記載の音声認識システム。

【請求項10】 前記設定手段は、前記分析手段で得られた特微量パラメータに基づいて、前記スカラ量子化のためのクラスタリングを変更させることを特徴とする請求項7乃至9のいずれかに記載の音声認識システム。

【請求項11】 音響情報を入力する入力工程と、

前記入力工程で入力した音響情報を分析して特微量パラメータを取得する分析工程と、

前記分析工程で得られた特微量パラメータに基づいて圧縮符号化のための処理条件を設定する設定工程と、

前記分析工程で得られた特微量パラメータを前記処理条件に従って圧縮符号化する変換工程と、

前記設定工程で設定された処理条件と、前記変換工程で圧縮符号化された特微量パラメータとに基づいて音声認識をする認識工程と  
を備えることを特徴とする音声認識方法。

【請求項12】 前記分析工程と前記設定工程と前記変換工程とを有する第1の装置と、前記認識工程を有する第2の装置とで構成され、

前記設定工程で設定された設定条件と前記変換工程で取得されたデータを前記第1の装置から前記第2の装置へ通知する通知工程を更に備えることを特徴とする請求項11に記載の音声認識方法。

【請求項13】 前記認識工程は、

前記処理条件を参照して前記圧縮符号化された特微量パラメータを復号する復号工程を備え、

前記復号工程で復号された特微量パラメータに基づいて音声認識処理を実行することを特徴とする請求項11または12に記載の音声認識方法。

【請求項14】 前記第2の装置が、前記通知工程で通知された前記処理条件をメモリに保持する保持工程を更に備えることを特徴とする請求項12に記載の音声認識方法。

【請求項15】 前記認識工程は、

前記処理条件と音響モデルに基づいて音声認識に関わる尤度計算の一部を行なう計算工程と、

前記計算工程による計算結果を用いて、前記変換工程で取得されたデータに対する尤度計算を行ない音声認識結果を得る工程とを備えることを特徴とする請求項11または12に記載の音声認識方法。

【請求項16】 前記計算工程で算出された計算結果をメモリに保持する保持工程を更に備えることを特徴とする請求項15に記載の音声認識方法。

【請求項17】 前記変換工程は、前記分析工程により得られた多次元音声パラメータを次元毎にスカラ量子化することを特徴とする請求項11乃至16のいずれかに記載の音声認識方法。

【請求項18】 前記スカラ量子化において、LBGアルゴリズムを用いることを特徴とする請求項17に記載の音声認識方法。

【請求項19】 前記スカラ量子化において、量子化対象データがガウス分布することを仮定し、量子化ステップが該分布で等確率となるよう量子することを特徴とする請求項17に記載の音声認識方法。

【請求項20】 前記設定工程は、前記分析工程で得られた特微量パラメータに基づいて、前記スカラ量子化のためのクラスタリングを変更させることを特

徴とする請求項17乃至19のいずれかに記載の音声認識方法。

【請求項21】 音響情報を入力する入力手段と、

前記入力手段で入力した音響情報を分析して特徴量パラメータを取得する分析手段と、

前記分析手段で得られた特徴量パラメータに基づいて圧縮符号化のための処理条件を設定する設定手段と、

前記設定手段で設定された処理条件を外部装置に通知する第1通知手段と、

前記分析手段で得られた特徴量パラメータを前記処理条件に従って圧縮符号化する変換手段と、

前記変換手段で得られたデータを前記外部装置に通知する第2通知手段と

を備えることを特徴とする情報処理装置。

【請求項22】 外部装置より圧縮符号化に係る処理条件を受信する第1受信手段と、

前記第1受信手段で受信した処理条件をメモリに保持させる保持手段と、

前記外部装置より圧縮符号化されたデータを受信する第2受信手段と、

前記保持手段に保持された処理条件を用いて前記第2受信手段で受信したデータに対する音声認識を実行する認識手段と

を備えることを特徴とする情報処理装置。

【請求項23】 音響情報を入力する入力工程と、

前記入力工程で入力した音響情報を分析して特徴量パラメータを取得する分析工程と、

前記分析工程で得られた特徴量パラメータに基づいて圧縮符号化のための処理条件を設定する設定工程と、

前記設定工程で設定された処理条件を外部装置に通知する第1通知工程と、

前記分析工程で得られた特徴量パラメータを前記処理条件に従って圧縮符号化する変換工程と、

前記変換工程で得られたデータを前記外部装置に通知する第2通知工程と

を備えることを特徴とする情報処理方法。

【請求項24】 外部装置より圧縮符号化に係る処理条件を受信する第1受

信工程と、

前記第1受信工程で受信した処理条件をメモリに保持させる保持工程と、

前記外部装置より圧縮符号化されたデータを受信する第2受信工程と、

前記保持工程に保持された処理条件を用いて前記第2受信工程で受信したデータに対する音声認識を実行する認識工程と

を備えることを特徴とする情報処理方法。

【請求項25】 請求項11乃至20のいずれかに記載の音声認識方法をコンピュータによって実現するためのプログラム。

【請求項26】 請求項23または24のいずれかに記載の情報処理方法をコンピュータによって実現するためのプログラム。

【請求項27】 請求項25または26のいずれかに記載のプログラムを格納する記憶媒体。

#### 【発明の詳細な説明】

##### 【0001】

##### 【発明の属する技術分野】

本発明は、音声認識システム及び装置とそれらの方法に関する。

##### 【0002】

##### 【従来の技術】

近年、音声認識技術の進歩に伴い、機器の入力インターフェースとして利用する試みがなされている。入力インターフェースとして音声技術を用いる場合は、当該機器に音声処理のための構成を導入し、当該機器内で音声認識を行ない、これを当該機器への入力操作として扱うことが一般的である。

##### 【0003】

一方、近年の小型携帯端末の発達は、小型携帯端末が多くの処理を実現することを可能とした。小型携帯端末のサイズ上の制約から充分な入力キーを具備することができない。このため、多様な機能を実現させるための操作指示に上記のような音声認識技術を利用したいという要望が高まっている。

##### 【0004】

これを実現する一つの方法としては小型携帯端末自身に音声認識エンジンを搭

載するという方法がある。しかし、このような小型携帯端末ではメモリやC P Uなどのリソースが限られていることが多い、高性能な認識エンジンを搭載することができない場合がある。そこで、小型携帯端末を無線などのネットワークでサーバと接続し、音声認識処理のうち処理コストの少ない部分は端末上で行ない、処理量が多い部分をサーバで行なうクライアント・サーバ型の音声認識が提案されている。

## 【0005】

この場合において、端末からサーバへ伝送するデータ量は少ない方が望ましいので、データを圧縮（符号化）してから伝送するのが一般的である。またそのための符号化方法に関して、携帯電話で使われているような一般的な音声符号化方法ではなく、音声認識に関するデータを送るのに適した符号化方法も提案されている。

## 【0006】

## 【発明が解決しようとする課題】

上述した、クライアント・サーバ型の音声認識において用いられる、音声認識に適した符号化においては、一般的には音声の特徴パラメータを求めた後、これらパラメータをスカラ量子化やベクトル量子化、あるいはサブバンド量子化により符号化する方法がとられる。そして、この場合の符号化は、音声認識する際の音響的特徴については特に考慮せざなされる。

## 【0007】

しかしながら、雑音環境下で音声認識を使用したり、音声認識の際使用するマイクの特性が一般的なものと異なる場合では、最適な符号化処理も異なってくる。例えば、上記の方法の場合、雑音環境下における音声の特徴パラメータの分布は、静謐な環境下における音声の特徴パラメータの分布とは違ってくるので、量子化の範囲もそれに適応していくことが望ましい。

## 【0008】

従来では、以上のような音響的特徴の変化を考慮にいれずに符号化を行なっていたため、雑音環境下などでは、認識率の劣化や、符号化の際の圧縮率を大きくできないという課題があった。

【0009】

本発明は上記の課題に鑑みてなされたものであり、音響的特徴の変化に応じた適切な符号化を可能とし、環境音の変化による認識率の低下や、符号化における圧縮率の低下を防止することを目的とする。

【0010】

【課題を解決するための手段】

上記の目的を達成するための本発明による音声認識システムは以下の構成を備える。すなわち、

音響情報を入力する入力手段と、

前記入力手段で入力した音響情報を分析して特微量パラメータを取得する分析手段と、

前記分析手段で得られた特微量パラメータに基づいて圧縮符号化のための処理条件を設定する設定手段と、

前記分析手段で得られた特微量パラメータを前記処理条件に従って圧縮符号化する変換手段と、

前記設定手段で設定された処理条件と、前記変換手段で圧縮符号化された特微量パラメータとに基づいて音声認識をする認識手段とを備える。

【0011】

また、上記の目的を達成するための本発明による音声認識方法は以下の構成を備える。すなわち、

音響情報を入力する入力工程と、

前記入力工程で入力した音響情報を分析して特微量パラメータを取得する分析工程と、

前記分析工程で得られた特微量パラメータに基づいて圧縮符号化のための処理条件を設定する設定工程と、

前記分析工程で得られた特微量パラメータを前記処理条件に従って圧縮符号化する変換工程と、

前記設定工程で設定された処理条件と、前記変換工程で圧縮符号化された特微量パラメータとに基づいて音声認識をする認識工程とを備える。

【0012】

また、上記の目的を達成するための本発明の情報処理装置は、音響情報を入力する入力手段と、前記入力手段で入力した音響情報を分析して特徴量パラメータを取得する分析手段と、

前記分析手段で得られた特徴量パラメータに基づいて圧縮符号化のための処理条件を設定する設定手段と、

前記設定手段で設定された処理条件を外部装置に通知する第1通知手段と、

前記分析手段で得られた特徴量パラメータを前記処理条件に従って圧縮符号化する変換手段と、

前記変換手段で得られたデータを前記外部装置に通知する第2通知手段とを備える。

【0013】

また、上記の目的を達成するための情報処理装置は、外部装置より圧縮符号化に係る処理条件を受信する第1受信手段と、前記第1受信手段で受信した処理条件をメモリに保持させる保持手段と、前記外部装置より圧縮符号化されたデータを受信する第2受信手段と、前記保持手段に保持された処理条件を用いて前記第2受信手段で受信したデータに対する音声認識を実行する認識手段とを備える。

【0014】

また、上記の目的を達成するための情報処理方法は、音響情報を入力する入力工程と、前記入力工程で入力した音響情報を分析して特徴量パラメータを取得する分析工程と、

前記分析工程で得られた特徴量パラメータに基づいて圧縮符号化のための処理条件を設定する設定工程と、

前記設定工程で設定された処理条件を外部装置に通知する第1通知工程と、

前記分析工程で得られた特徴量パラメータを前記処理条件に従って圧縮符号化する変換工程と、

前記変換工程で得られたデータを前記外部装置に通知する第2通知工程とを備える。

## 【0015】

更に、上記の目的を達成するための情報処理方法は、  
外部装置より圧縮符号化に係る処理条件を受信する第1受信工程と、  
前記第1受信工程で受信した処理条件をメモリに保持させる保持工程と、  
前記外部装置より圧縮符号化されたデータを受信する第2受信工程と、  
前記保持工程に保持された処理条件を用いて前記第2受信工程で受信したデータに対する音声認識を実行する認識工程とを備える。

## 【0016】

## 【発明の実施の形態】

以下、本発明の好適な実施の形態について、添付図面を参照して説明する。

## 【0017】

## &lt;第1実施形態&gt;

図1は、第1実施形態に係る音声認識システムの構成を示したブロック図である。また図2、図3は図1の構成図に記載された音声認識システムの動作を説明するフローチャートである。以下、動作例を含め、図1と図2、図3を関連させて説明する。

## 【0018】

図1において、100は端末部であり、携帯電話等を含む種々の携帯端末が適用可能である。101は音声入力部であり、マイク等によって音声信号を取り込み、これをデジタル化する。102は音響処理部であり、音響分析を行なって、多次元音響パラメータを生成する。なお、音響分析には、メルケプストラムやデルタメルケプストラムなど、一般的に音声認識に用いられる分析法を使用できる。103は処理切換部であり、図2及び図3を参照して後述する初期設定時の処理及び音声認識時の処理とでデータの流れを切換える。

## 【0019】

104は音声通信情報生成部であり、音響処理部102で得られた音響パラメータに対して符号化を行なうためのデータの生成を行なう。本実施形態では、音

声通信情報生成部104は、音響パラメータの各次元のデータを任意のクラス（本例では16stepとする）にクラスタリングし、クラスタリングにより分割した結果を用いてクラスタリング結果テーブルを生成する。105は音声通信情報保持部であり、音声通信情報生成部104で生成したクラスタリング結果テーブルを保持する。なお、音声通信情報保持部105においてクラスタリング結果テーブルを保持させる記録媒体としては、RAM等のメモリ、フロッピーディスク（FD）、ハードディスク（HD）など種々のものが利用できる。

## 【0020】

106は符号化部であり、音声通信保持部105に記録されたクラスタリング結果テーブルを用いて音響処理部102より得られた多次元音響パラメータを符号化する。107は通信制御部であり、クラスタリング結果テーブルや符号化された音響パラメータ等を通信回線300に対して送出する。

## 【0021】

200はサーバ部であり、端末部100より送られてくる、符号化された多次元音響パラメータについて音声認識を行なう。サーバ部200は通常のパソコン等によって構成することができる。

## 【0022】

201は通信制御部であり、回線300を介して端末部100の通信制御部107より送信されたデータを受信する。202は処理切換部であり、図2及び図3を参照して後述する初期設定時の処理及び音声認識時の処理とでデータの流れを切換える。

## 【0023】

203は音声通信情報保持部であり、端末部100より受信したクラスタリング結果テーブルを保持する。なお、音声通信情報保持部203においてクラスタリング結果テーブルを保持させるための記録媒体としては、RAM等のメモリ、フロッピーディスク（FD）、ハードディスク（HD）など種々のものが利用できる。

## 【0024】

204は復号化部であり、音声通信情報保持部203に保持されたクラスタリ

ング結果テーブルを参照して、通信制御部201で端末部100より受信された符号化データ（多次元音響パラメータ）を復号する。205は音声認識部であり、音響モデル保持部206に保持された音響モデルを用いて、復号化部204で得られた多次元音響パラメータについて認識処理を行なう。

## 【0025】

207はアプリケーションであり、音声認識の結果に基づいて各種処理を実行する。アプリケーション207はサーバ部200側で実行されるものであっても、端末部100側で実行されるものであってもよい。但し、端末部100側で実行されるアプリケーションの場合は、音声の認識結果を通信制御部201、107を介して端末部100へ通知する必要がある。

## 【0026】

なお、端末部100の処理切換部103は初期設定時には音声通信情報生成部104へ、音声認識処理時には符号化部106へデータが流れるように接続を切換える。同様に、サーバ部200の処理切換部202は初期設定時には音声通信情報保持部203へ、音声認識処理時には復号化部204へデータが流れるように接続を切換える。これら処理切換部103と202は連動して動作する。その切換えは、例えば初期学習モードと認識モードの二種類のモードがあり、ユーザの指示によって認識使用前に学習を行うべく初期学習モードが指示された時は処理切換部103は音声通信情報生成部104へ、処理切換部202は音声通信情報保持部203へデータが流れるように接続を切換える。実際に認識するときはユーザが認識モードを指示するので、それに応答して処理切換部103は符号化部106へ、処理切換部202は復号化部204へデータが流れるように接続を切換える。

## 【0027】

なお、300は端末部100とサーバ部200を接続する通信回線であり、有線、無線にかかわらずデータを転送できる種々の通信手段が利用可能である。

なお、上述した端末部100、サーバ部200の各部は、それぞれが有するCPUがメモリに格納された制御プログラムを実行することによって実現されるものである。もちろん、各部の一部或いは全ての機能をハードウェアによって実現

してもよい。

#### 【0028】

以下、図2及び図3のフローチャートにより、上記音声認識システムにおける動作を詳細に説明する。

#### 【0029】

まず音声認識開始前に、図2のフローチャートに示される初期設定を行なう。

初期設定では符号化データを音響環境へ適応させるための符号化条件の設定を行なう。この初期設定は行なわなくとも、例えば静かな環境での音響状態を基に作成された既定値を使用することにより、音声データの符号化及び音声認識は可能であるが、初期設定を行なうことにより認識率の向上が見込まれる。

#### 【0030】

初期設定では、まずステップS2において、音声入力部101で音響データを取り込み、さらにA/D変換を行なう。入力される音響データは実際に使用される音環境またはそれに類似した音環境で音声を発声した場合の音声データである。この音響データには使用したマイクロフォンの特性の影響も反映される。また背景や機器内部で発生する雑音がある場合、その影響も受けたデータになる。

#### 【0031】

ステップS3では、音響処理部102において、音声入力部101で入力された音響データの音響分析を行なう。上述したように、音響分析はメルケプストラムやデルタメルケプストラムなど、一般的に音声認識に用いられる分析法を使用できる。上述のように、処理切換部103は初期設定時には音声通信情報生成部104の側と接続するので、ステップS4では、音声通信情報生成部104により符号化処理のためのデータの作成が行なわれる。

#### 【0032】

この音声通信情報生成部104で行われるデータの作成方法について以下説明する。音声認識用の符号化に関しては、音響パラメータを求めそれをスカラ量子化、ベクトル量子化、サブバンド量子化する方法などが用いられる。本実施形態においては、とくにその手法を限定する必要はなく、いずれの方法も使用できるが、ここではスカラ量子化による方法について説明する。この方法ではステップ

S3の音響分析によって求められた多次元音響パラメータの各次元をスカラ量子化する。スカラ量子化には様々な方法が可能である。

#### 【0033】

以下に2つの例を示す。

##### 1) LBGによる方法:

クラスタリング手法として一般的に用いられる LBG 法を使用する。音響パラメータの各次元のデータを LBG 法を用いて任意のクラス（例えば 16 step）に分割する。

##### 2) モデルを仮定する方法:

音響パラメータの各次元のデータが、例えばガウス分布に従うと仮定する。各次元の分布全体の 3σ の範囲内を面積等分割、つまり等確率になるよう分割し、例えば 16 step にクラスタリングする。

#### 【0034】

さらにステップ S6において、音声通信情報生成部 104 で求められたクラスタリング結果テーブルをサーバ部 200 に転送する。転送に当たっては、端末部 100 における通信制御部 107、通信回線 300、およびサーバ部 200 の通信制御部 201 を用い、サーバ部 200 にクラスタリング結果テーブルが転送される。

#### 【0035】

サーバ部 200 では、ステップ S7において、通信制御部 201 によってクラスタリング結果テーブルを受信する。このとき、処理切換部 202 は音声通信情報保持部 203 と通信制御部 201 とを接続しており、ステップ S8 にて、音声通信情報保持部 203 に受信されたクラスタリング結果テーブルが記録されることになる。

#### 【0036】

次に音声認識時の処理について説明する。図3は音声認識時の処理の流れを示したフローチャートである。

#### 【0037】

音声認識では、まずステップ S21において、音声入力部 101 で認識対象の

音声を取り込み、A/D変換を行なう。ステップS22では音響処理部102で音響分析を行なう。音響分析は、初期設定時と同様にメルケプストラムやデルタメルケプストラムなど、一般的に音声認識に用いられる分析法を使用できる。音声認識時では、処理切換部103は音響処理部102と符号化部106を接続する。従って、ステップS23において、符号化部106は音声通信保持部105に記録されたクラスタリング結果テーブルを用いて、ステップS22で求めた多次元特徴量パラメータを符号化する。すなわち、符号化部106で、次元毎のスカラ量子化を行なう。

#### 【0038】

符号化により各次元のデータはたとえば4bit(16step)のデータに変換される。例えばパラメータの次元数が13次元、各次元のデータが4bitで、分析周期が10msすなわち、毎秒100フレームのデータが転送される場合のデータ量は、

$$13 \text{ (次元)} \times 4 \text{ (bit)} \times 100 \text{ (frame/s)} = 5.2 \text{ kbps}$$

となる。

#### 【0039】

次にステップS24、S25で符号化データの送出および受信を行なう。データ転送に当たっては、上述のように端末部100における通信制御部107、通信回線300、およびサーバ部200の通信制御部201を用いる。通信回線300では有線、無線にかかわらずデータを転送できる種々の通信手段が利用可能である。

#### 【0040】

音声認識処理時では、処理切換部202は通信制御部201と復号化部204とを接続する。よって、ステップS26で復号化部204が音声通信情報保持部203に記録されたクラスタリング結果テーブルを用いて、通信制御部201で受信した多次元特徴量パラメータを復号する。この復号化により音響パラメータが得られる。ステップS27では、ステップS26で復号されたパラメータを用いて音声認識を行なう。この音声認識は、音声認識部205が、音響モデル保持部206に保持された音響モデルを用いて行う。但し、一般の音声認識とは異なる

り、音響処理部はない。これは、復号化部204で復号するデータが音響パラメータそのものだからである。また音響モデルとしては例えばHMM (Hidden Markov Model) が用いられる。ステップS28では、ステップS27の音声認識によって得られた音声認識結果を用いてアプリケーション207を動作させる。アプリケーション207はサーバ部200にあっても、端末部100にあっても、あるいは両方に分散されていても構わない。アプリケーション207が端末部100にある場合や分散されている場合は、通信制御部107、201および通信回線300を使用し、認識結果やアプリケーションの内部状態のデータ等を転送する必要がある。

#### 【0041】

以上説明したように、第1実施形態によれば、音響パラメータの符号化処理において、音響状態に適応した符号化を行なうためテーブル（クラスタリング結果テーブル）を用いて符号化及び復号化が行なわれるので、音響的特徴の変化に対して適切な符号化が可能となる。このため、環境音の変化による認識率の低下を防止できる。

#### 【0042】

##### <第2実施形態>

第1実施形態では、音響状態に適応化した符号化条件（クラスタリング結果テーブル）を作成し、この符号化条件を符号化部106と復号化部204の間で共有して符号化処理復号化処理を実行し、適切な音声データの送信、音声認識処理を実現した。第2実施形態では、更に処理の高速化のために、符号化したデータを復号化せずに認識を行なう方法について説明する。

#### 【0043】

図4は、第2実施形態に係る音声認識システムの構成を示したブロック図である。また図5、図6は図4の構成図に記載された音声認識システムの動作を説明するフローチャートである。以下、動作例を含め、図4と図5、図6を関連させて説明する。

図4において、第1実施形態で示した構成と同様のものには、同一の参照番号を付してある。同図からわかるように、端末部100は第1実施形態の構成と同

じである。サーバ部500において、処理切換部502は初期設定時には通信制御部201と尤度情報生成部503とを接続し、音声認識処理時には通信制御部201と音声認識部505とを接続する。

#### 【0044】

503は尤度情報生成部であり、入力されたクラスタリング結果テーブルと音響モデル保持部506に保持された音響テーブルに基づいて尤度情報を生成する。ここで生成される尤度情報により符号化データを復号せずに音声認識することができる。尤度情報とその生成については後述する。504は尤度情報保持部であり、誘導情報生成部503で生成された尤度情報を保持する。なお、尤度情報保持部504において尤度情報を保持させるための記録媒体としては、RAM等のメモリ、フロッピーディスク(FD)、ハードディスク(HD)など種々のものが利用できる。

#### 【0045】

505は音声認識部であり、尤度計算部508及び言語検索部509を備える。音声認識部505の動作については後述するが、通信制御部201より入力された符号化データに対して、尤度情報保持部504に保持された尤度情報を用いて認識処理を実行する。

#### 【0046】

以下、図5及び図6を参照して第2実施形態による音声認識処理を説明する。

#### 【0047】

まず、音声認識開始前に初期設定を行なう。第1実施形態と同様に、初期設定は符号化データの音響環境への適応を行なうためのものである。この初期設定は行なわなくとも符号化データに関し既定値を使用することにより音声認識は可能であるが、初期設定を行なうことにより認識率の向上が見込まれる。

#### 【0048】

端末部100におけるステップS40～S45の各処理については第1実施形態(ステップS1～S6)と全く同一であるため説明を省略する。以下、サーバ部500の初期設定処理について説明する。

#### 【0049】

まずステップS46では、端末部100で生成された音声通信情報（本例ではクラスタリング結果テーブル）を通信制御部501で受信する。処理切換部502は初期設定処理では尤度情報生成部503側と接続している。よって、ステップS47において尤度情報が生成される。以下、尤度情報の生成について述べる。尤度情報の生成は尤度情報生成部503において、音響モデル保持部506に保持されている音響モデルを用いて行われる。この音響モデルは例えばHMMなどで表現される。

#### 【0050】

尤度情報の生成方法には種々のものがあるが、ここではスカラ量子化を利用する方法について述べる。第1実施形態で説明したように、ステップS40～S45の端末部100の処理で、多次元音響パラメータの次元ごとのスカラ量子化用のクラスタリング結果テーブルが得られる。このテーブルに保持される各量子化点の値および音響モデルにより、各量子化点に関して尤度計算の一部を行なう。この値を尤度情報保持部504に保持する。認識時には符号化データとして受信されるスカラ量子化値に基づき、テーブル参照により尤度計算を行なうため復号化が不要になる。

#### 【0051】

このようなテーブル参照による尤度計算の方法は、文献、嵯峨山らによる「音声認識における新しい高速化」、平成8年度春季日本音響学会講演論文集1-5-12に詳しい。またスカラ量子化の他ベクトル量子化法や、各次元で混合分布演算を予めして加算を省略する方法なども使用できる。これらについても上記文献に紹介されている。上記計算結果は、ステップS48において、尤度保持部504にスカラ量子化値に対するテーブルの形で保持される。

#### 【0052】

次に図6を参照して、第2実施形態による音声認識処理の流れについて説明する。端末部100でのステップS60～S64に示される処理は、第1実施形態（ステップS20～S24）と同一であるため説明を省略する。

#### 【0053】

ステップS65では、ステップS20～S24の処理で求められた多次元音響

パラメータの符号化データを、サーバ部500の通信制御部501で受信する。音声認識時においては、処理切換部502は尤度計算部508側と接続される。音声認識部505は、尤度計算部508および言語探索部509に分けて表現が可能である。ステップS66では、尤度計算部508により尤度を計算するが、この場合、音響モデルではなく、尤度情報保持部504に保持されたデータを用い、スカラ量子化値に対するテーブル参照を行なって計算する。計算の詳細は上記の文献に詳しく記載されているので、ここでは説明を省略する。

## 【0054】

ステップS67では、ステップS66による尤度計算の結果に対し、言語探索を行なって認識結果を得る。言語探索は、単語辞書、ネットワーク文法、n-gramのような言語モデル等、一般的に音声認識に使われる文法を用いて行なう。ステップS68では、得られた認識結果を用いアプリケーション507を動作させる。なお、第1実施形態と同様に、アプリケーション507はサーバ部500にあっても、端末部100にあっても、あるいは両方に分散されていても構わない。アプリケーションが端末部100にある場合や分散されている場合は、通信制御部407、501および通信回線300を使用し、認識結果やアプリケーションの内部状態のデータ等を転送する必要がある。

## 【0055】

以上のように、第2の実施形態によれば、符号化したデータを復号化することなく音声認識できるので、処理の高速化を図ることができる。

## 【0056】

以上説明した第1及び第2実施形態の音声認識処理は、音声認識を利用したアプリケーション一般について利用できる。特に、端末部100として小型携帯端末を用い、機器の制御や情報検索を音声入力により行なうような場合に適している。

## 【0057】

また、上記各実施形態によれば、音声認識用の符号化を用いて、音声認識処理を機器に分散させて動作させる場合において、符号化処理が背景雑音や内部雑音、またはマイクロフォンの特性などに応じて行われることになる。このため雑音

環境下やマイクロフォンの特性が異なる場合などにおいても、認識率の劣化を防止できるとともに、効率のよい符号化が可能となり、通信路の伝送データ量を押さえることができるなどのメリットが得られる。

## 【0058】

なお、本発明の目的は、前述した実施形態の機能を実現するソフトウェアのプログラムコードを記録した記憶媒体を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ（またはC P UやM P U）が記憶媒体に格納されたプログラムコードを読み出し実行することによっても、達成されることは言うまでもない。

## 【0059】

この場合、記憶媒体から読み出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコードを記憶した記憶媒体は本発明を構成することになる。

## 【0060】

プログラムコードを供給するための記憶媒体としては、例えば、フロッピディスク、ハードディスク、光ディスク、光磁気ディスク、C D - R O M、C D - R、磁気テープ、不揮発性のメモリカード、R O Mなどを用いることができる。

## 【0061】

また、コンピュータが読み出したプログラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているO S（オペレーティングシステム）などが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

## 【0062】

さらに、記憶媒体から読み出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるC P Uなどが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言

うまでもない。

【0063】

【発明の効果】

以上説明したように、本発明によれば、音響的特徴の変化に応じた適切な符号化が可能となり、環境音の変化による認識率の低下や、符号化における圧縮率の低下を防止することができる。

【図面の簡単な説明】

【図1】

第1実施形態における音声認識システムの構成を示すブロック図である。

【図2】

第1実施形態における音声認識システムの初期設定を説明するフローチャートである。

【図3】

第1実施形態における音声認識システムの音声認識処理を説明するフローチャートである。

【図4】

第2実施形態における音声認識システムの構成を示すブロック図である。

【図5】

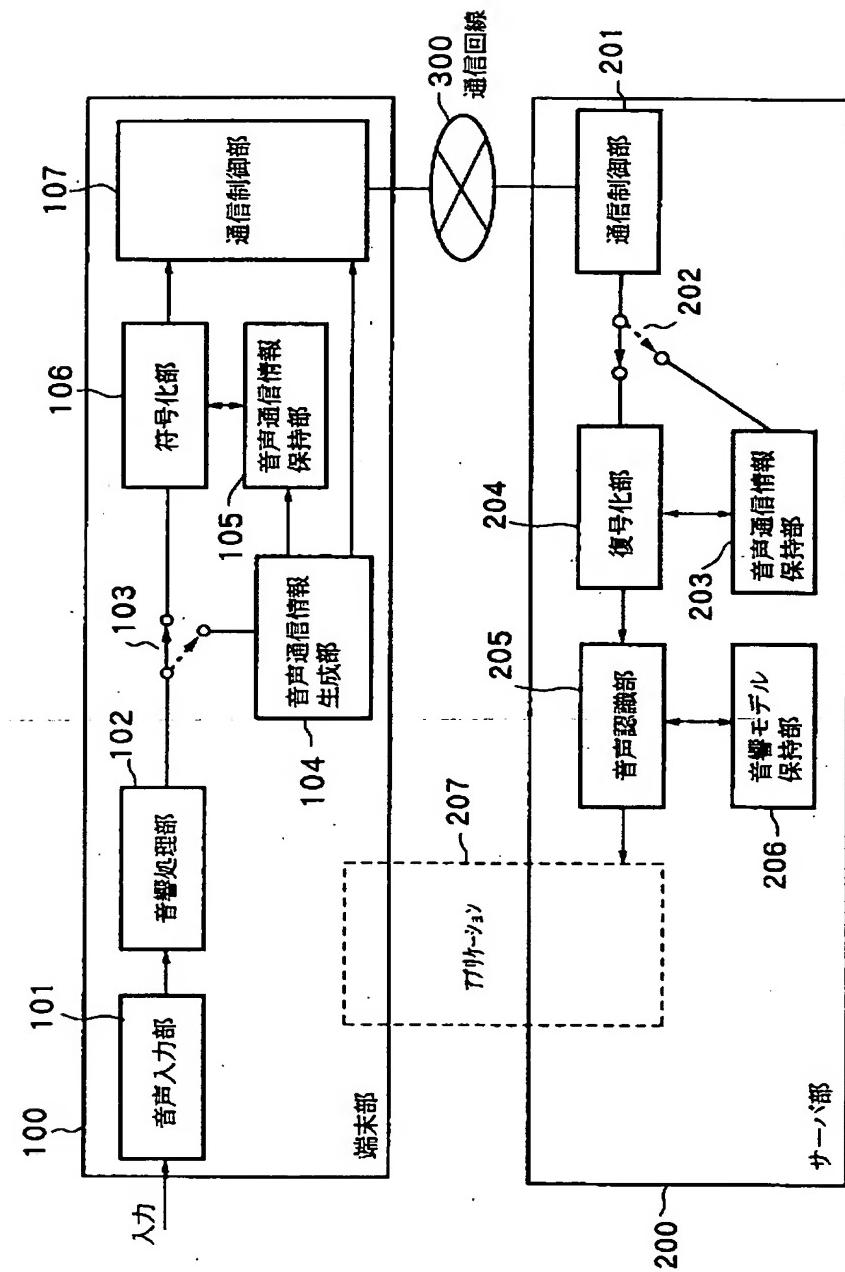
第2実施形態における音声認識システムの初期設定を説明するフローチャートである。

【図6】

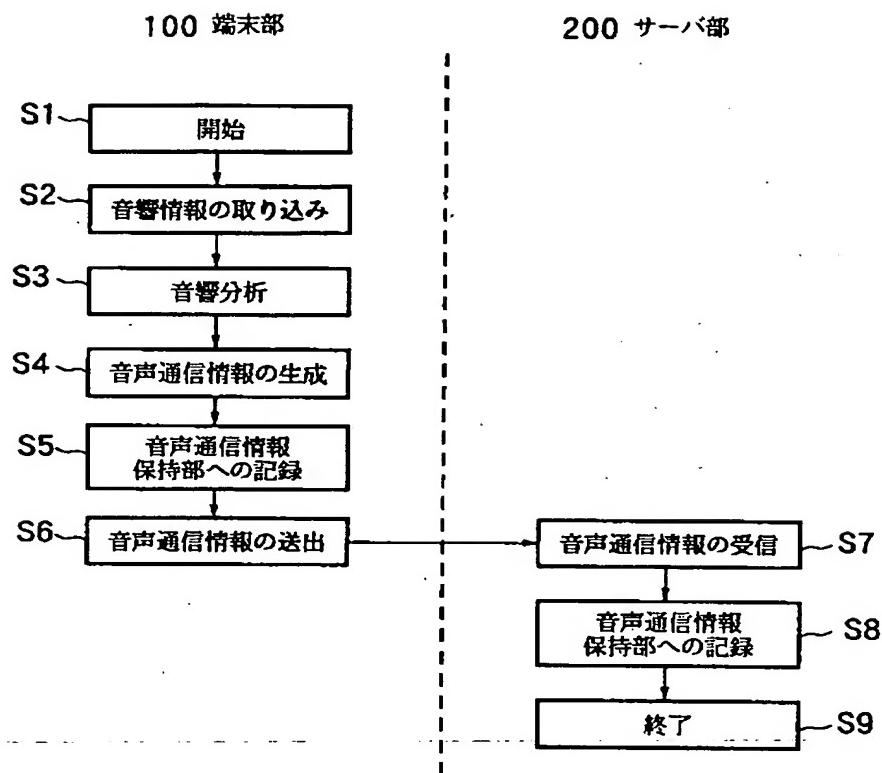
第2実施形態における音声認識システムの音声認識処理を説明するフローチャートである。

【書類名】 図面

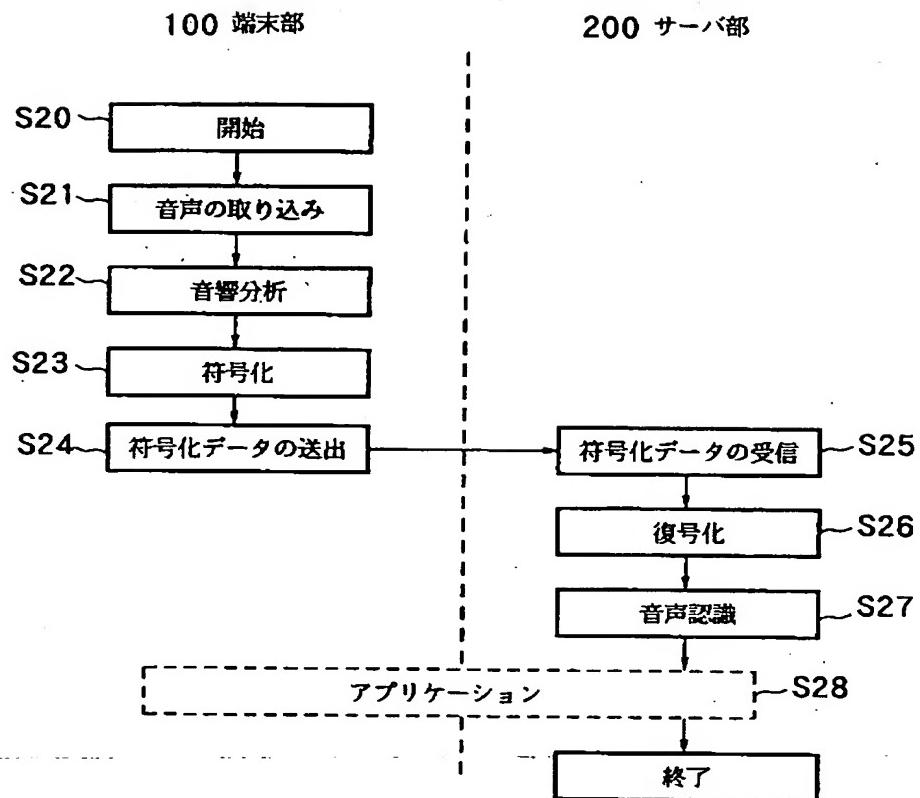
〔図1〕



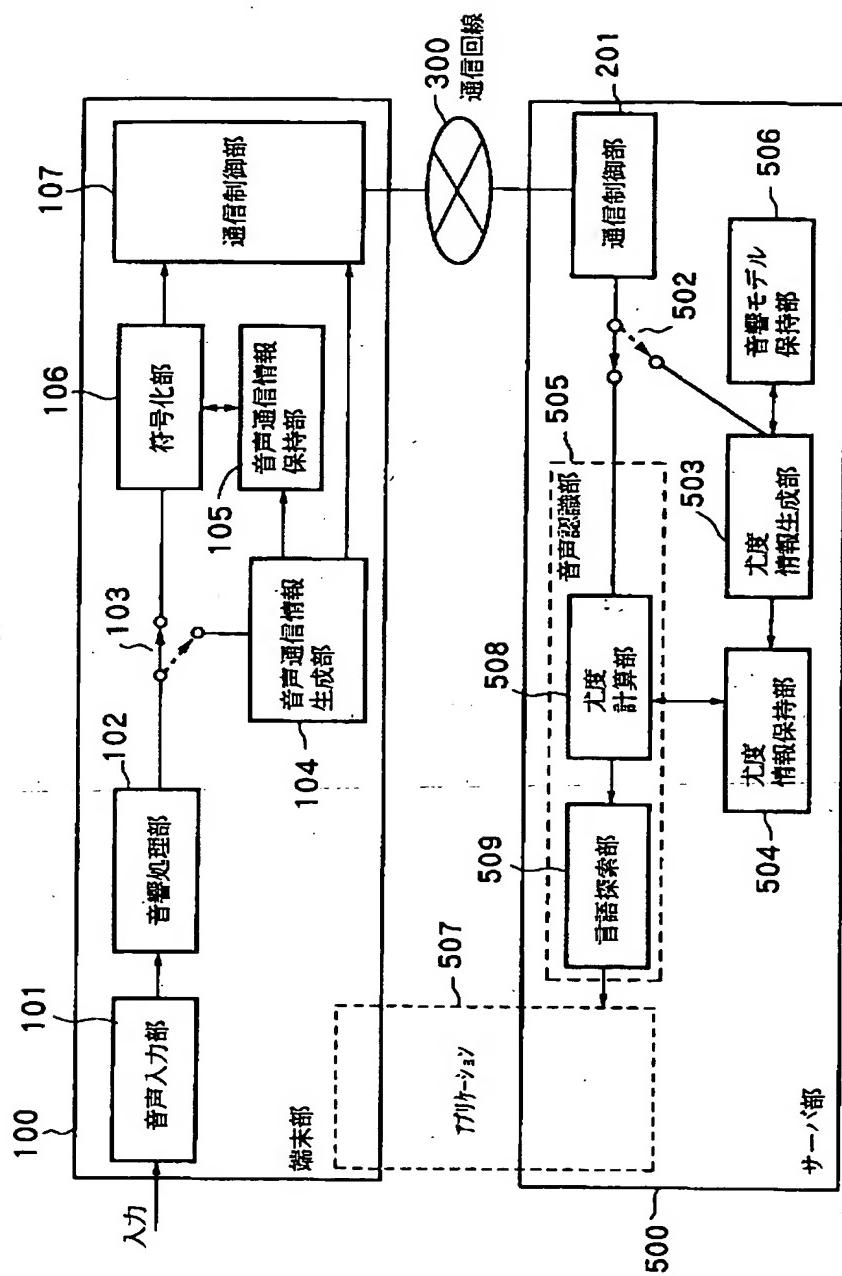
【図2】



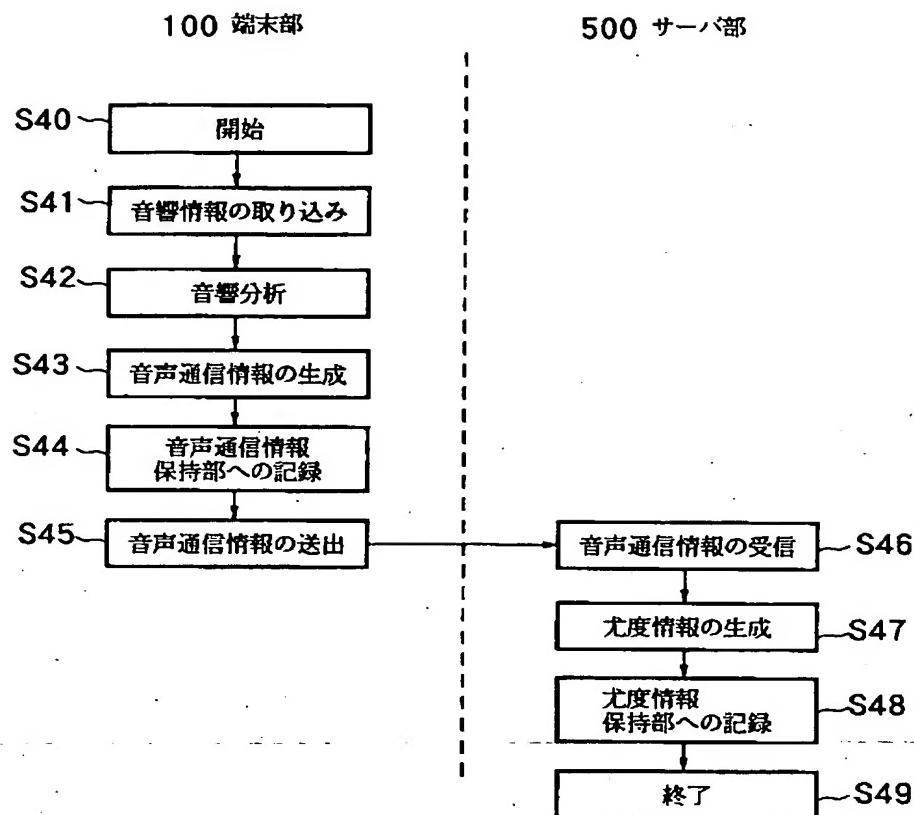
【図3】



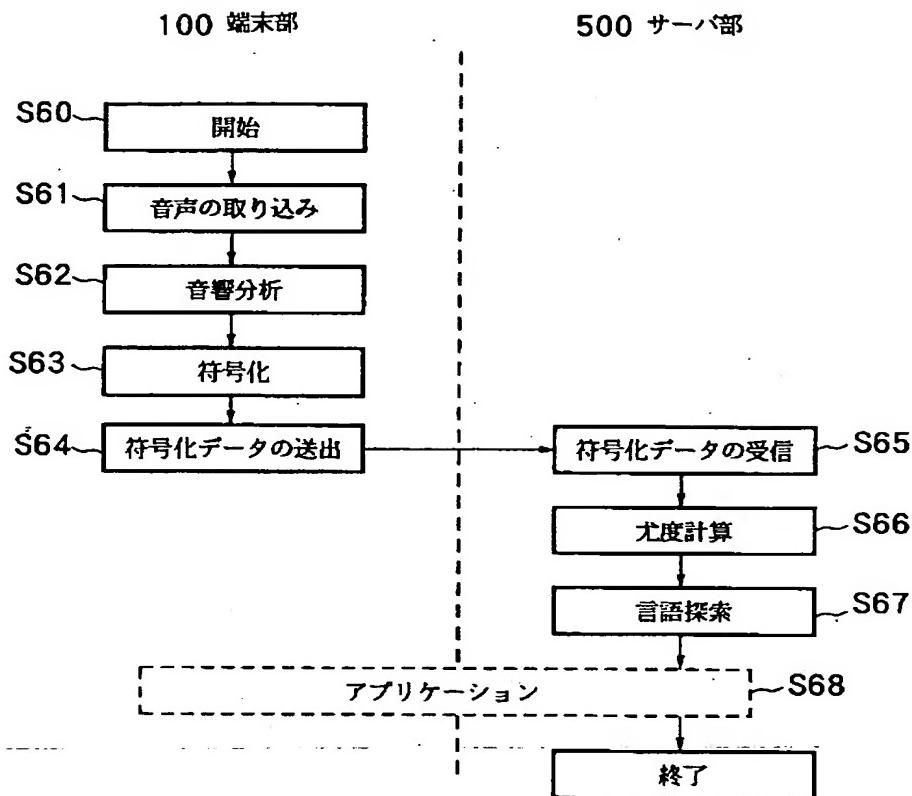
【図4】



【図5】



【図6】



【書類名】 要約書

【要約】

【課題】 音響的特徴の変化に応じた適切な符号化を可能とし、環境音の変化による認識率の低下や符号化における圧縮率の低下を防止する。

【解決手段】 端末部100において、音響入力部101で入力された音響情報を音響処理部102が分析し、多次元特徴量パラメータを取得する。初期設定時においては、音声通信情報生成部104がこの多次元特徴量パラメータに基づいて圧縮符号化のための処理条件を設定し、音声通信情報保持部105に保持するとともに、サーバ部200の音声通信情報保持部203にこの処理条件が保持される。音声認識においては、符号化部106が、音響処理部102で得られた特徴量パラメータを音声通信情報保持部105に保持された処理条件に従って圧縮符号化し、これをサーバ部200の復号化部204が、音声通信情報保持部203に保持された処理条件を用いて復号する。

【選択図】 図1

出願人履歴情報

識別番号 [000001007]

1. 変更年月日 1990年 8月30日

[変更理由] 新規登録

住 所 東京都大田区下丸子3丁目30番2号

氏 名 キヤノン株式会社